Info. System — Dr. Sherin — 14/2/017

Big Data Analysis

Big Data
- 1. A lot of data
  - data Complexity
- 2. Vareity of data
  - 1. Structured
  - 4. Unstructured # XML
  - 2. Semistructured
  - 3. Quasi Structured
    - web search history
    - web click stream
- 3. data streaming

test:

3V model:
Volume Vareity Velocity

Value of data ≡ Business value

4V model:
Volume Vareity Velocity Value

Certainity of data ≡ Veracity

5V model:
Volume Vareity Velocity Value Veracity

R lang. ≡ Tool

test: ch/c of Big data

Module 1

distribution of data →
timeliness → rate of data creation, every day/month/year/sec

data ~~Analy~~ scientist
- must be doubtful
- must be aware of lots of fields

OLAP → Online Analytical Processing

test slide 17

Data Island → not connected, easy, any one can make it not secured, replication may exist (redundancy), data can get lost

data warehouse → secure, Backup exists, connected
DB Adminstrator is in Control, No replication
- responsible for permissions

Sandbox → Analyst owns it not ADministrator
Big Size, not small
Variety of data, more accurate results.
shadow file sys. → I see the data in the Data warehouse but don't have a replication of it & can't change it without permission

mini-Case Study
requirements
- more security
- relations between ~~countries~~ currencies
- laws / legal requirements
- distributed storage

Controlled access to data from all Banks ....
- more services (ATM)(Employees)

requirements from Analyst point of view:
- Velocity
- Variety (critic rule)
- Volume
- Tools & methods to manage the data
- data driven (manages based on the data he gets)

KPI ≡ Key Performance Indicator

TP True +ve
FN False -ve
TN True -ve
FP False +ve

$\frac{TP+TN}{total}$ = accuracy

Business inteligence, data Analyst ... oral test
- Skills
- how can they help you

~~Business~~
Business Intelligence → Just an indicator
doesn't take decisions
uses KPI, deals with structured data

Data Analyst → ~~does~~ gives Analysis ?
gives decisions
High knowledge in AI, machine learning, Math, data mining